THE EARTH SCIENCE REMOTE ACCESS TOOL

Robert G. Raskin[*][+] and Elaine R. Dobinson[+]
Physical Oceanography Distributed Active Archive Center (PO.DAAC)
Jet Propulsion Laboratory
Pasadena, CA

1. INTRODUCTION

The Earth Science Remote Access Tool (ESRAT) is an http-based application that facilitates internet access to user-defined subsets of Earth Science data. It addresses the variety of data formats used in the Earth Sciences by providing a common data model with translators for the models inherent in many standard formats. It enables applications expecting local data in a prescribed format to access remote data in that or other formats. Data from multiple sites and in multiple formats can be combined readily into a single MATLAB plot.

A Java GUI is used to select spatial/temporal subsets of interest; subsetting is performed prior to data transmission, reducing bandwidth requirements. The tool is well suited for identification of and access to corroborating data in the calibration and validation of EOS instruments, and is initially being used in this application.

2. CLIENT-SERVER DESIGN

ESRAT features a client-server architecture depicted in Figure 1. It builds upon the client-server design of the Distributed Oceanographic Data System (DODS) (Gallagher and Milkowski,1995) developed at the University of Rhode Island and the Massachusetts Institute of Technology. ESRAT and DODS address the lack of common data models in the geophysical sciences by providing a common data model with translators for the data models inherent in many standard formats (HDF, netCDF, MATLAB, DSP and JGOFS).

The client side features a Java applet GUI that allows the user to interactively select a region and time interval of interest. The client establishes network connections with DODS servers and loads in-

---

Corresponding author address: Robert Raskin, Jet Propulsion Laboratory 525-389, Pasadena, CA 91109; e-mail: rob.raskin@jpl.nasa.gov.

coming data into a MATLAB session by invoking a web browser helper application.

The server side features master directory, dataset catalog, and data access servers. Each is implemented as a C++ CGI program. The master directory contains a list of dataset holdings at local or remote sites and their associated URLs.

The dataset catalogs contain the mapping of array rows/columns and file names to spatial/temporal coordinates. This information permits ESRAT/DODS clients to query by geographical and temporal coordinates and to return only the the desired subsets and variables. The catalogs need not reside at the same physical location as the datasets themselves.

The data access server translates the desired data into an intermediate DODS model made up of "Arrays", "Grids", and "Sequences" (tables). A DODS/HDF server was built as part of this effort and supports translation from the SDS, raster, and Vdata elements of the HDF data model. DODS clients perform a reverse translation, from the intermediate DODS model into a visualization/analysis package such as MATLAB.

The server performs spatial/temporal subsetting of the desired variables prior to data transmission. For Arrays and Grids, server support is provided for subsetting by row and column; for Sequences, subsetting also can be carried out by value of the dependent variables.

3. SAMPLE SESSION

In a typical session, the user accesses the GUI using a web browser. The GUI provides a list of available datasets and their temporal coverages (Figure 2). The user selects a region and time interval of interest, after which the dataset catalog server returns a graphical display of the subsets satisfying the search criteria. The variables of interest are selected at this point, and a final narrowing down of the subsets can be made before the data access server downloads the data. The data are loaded directly into a MATLAB script, making use of a browser helper program that directs the incoming data.

```
          GUI                              Master
        Select:          http://dods.jpl.nasa.gov/master.cat       catalog
        dataset,
        spatial/time
        range,           Interactive selection of subsets          Dataset
        variables                                                  catalogs

        Create
      constrained    http://sea.hawaii.edu/file.hdf?temp[0:54][360:380]   Data
          URL

       MATLAB                                                     Translate
                                                                 from HDF
                                                                 to DODS
                                                                  format

       Translate                                                   Extract
      from DODS    Temperature values within specified region/time   subset
        format
```

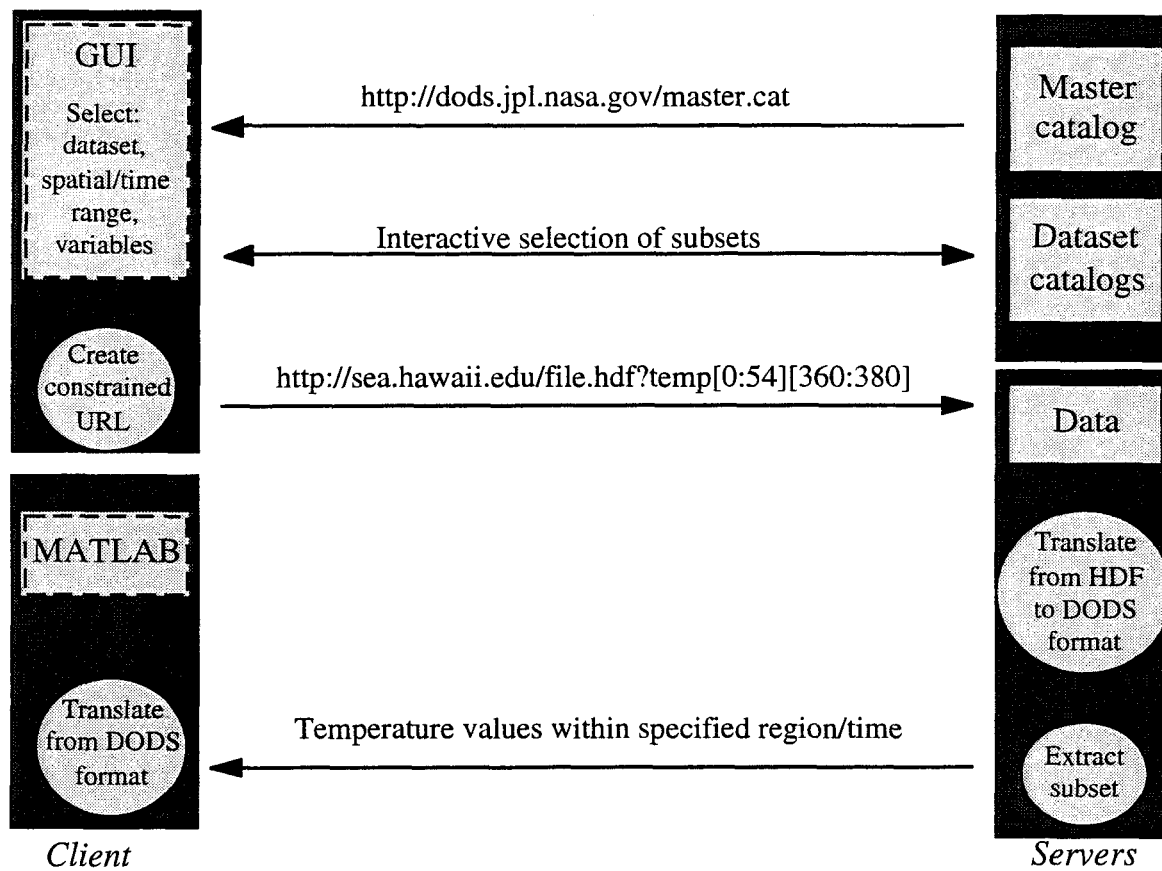Client   Servers

Figure 1. Client-server system architecture of ESRAT. Servers may be distributed over several sites.

4. APPLICATION

ESRAT has been applied to support the calibration and validation of the NASA scatterometer (NSCAT), an EOS instrument that measures winds over the global oceans. The satellite carrying NSCAT failed in June 1997, after nine months of successful data transmission; nevertheless, reprocessing of the data is still ongoing. To determine the accuracy of the data, NSCAT scientists require access to numerous ground-truth and satellite datasets coincident in space and time with the NSCAT data. Several of these datasets are included in the master catalog, including TAO buoy, NCEP model, and TOPEX/Poseidon. The coverage plotting capability of ESRAT allows the determination of overlapping data in space and time to aid in the discovery of corroborating data.

5. LESSONS LEARNED

One of the lessons learned in the development of the prototype was that 100% interoperability between formats is non-trivial. HDF data structures are richer than those in netCDF or in the intermediate DODS format. Some structural information may be lost in the conversion. In particular, our DODS server currently does not provide translation support for HDF Vgroups; nevertheless, useful applications of the tool still abound.

It also was learned that software technology is highly volatile. DODS was unique when it first appeared several years ago, enabling network connections for scientific data transfer. Now, standard software products are appearing on the market with the capability of accessing remote files as readily as local files.

6. ONGOING AND FUTURE WORK

We currently are converting our data catalogs to a commercial object-relational database management system (ORDBMS). This strategy will enable the system to scale efficiently to large data volumes typical of EOSDIS archives. The Java Database Connectivity (JDBC) package will replace the dataset catalog server as the interface to the dataset catalogs.

In the subsequent phase of the project, we will replace the data access server with Java-compliant tools such as RMI/CORBA. This will enable us to increase the number of supported clients and servers by leveraging off the work of vendors who are creating
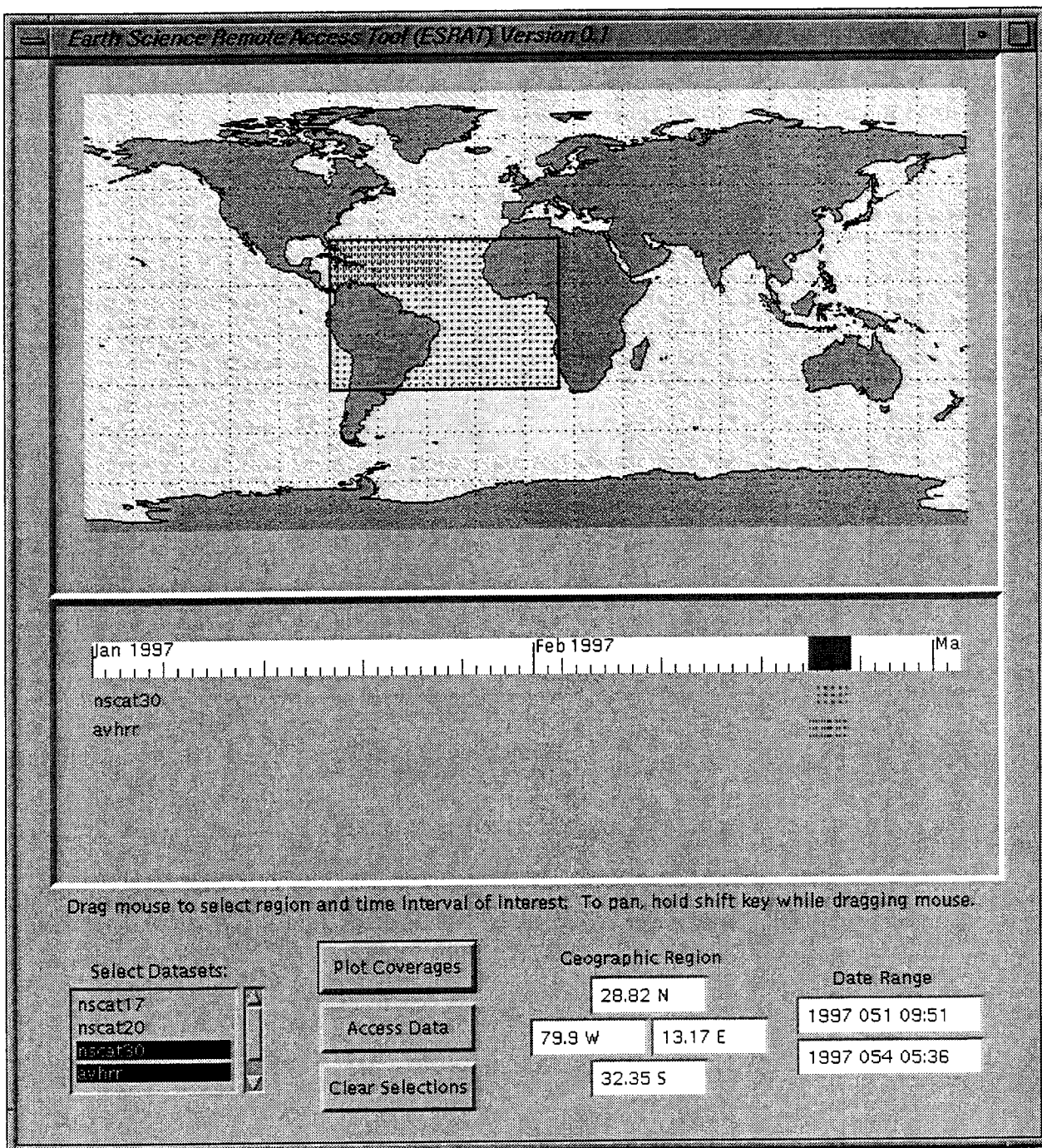
Figure 2. Sample ESRAT session displaying spatial/temporal coverages of two datasets (NSCAT Level 3.0 and AVHRR) within region of interest (tropical Atlantic) and time interval of interest (late February)

Java translators/interfaces. For example, NCSA recently released the Java HDF Interface (JHI) that translates HDF data structures to Java classes. Also, IDL is developing an analogous interface that translates Java classes to IDL that will enable the development of an IDL client. The platform independence of Java permits Java's data model (in terms of arrays, classes, and streams) to be used as the intermediate data model in this scheme.

Other plans are to provide support for the HDF-EOS format. Files in this format contain the spatial and temporal bounds embedded as metadata (using standard conventions adoptd by EOSDIS) to simplify the creation of dataset catalogs. HDF-EOS also extends HDF by providing three spatial data types

(grid, swath, point); however, no datasets are currently available that use these data types.

A White Paper describing our entire project in greater detail can be found at: http://dods.jpl.nasa.gov/wp.

7. ACKNOWLEDGMENTS

8. REFERENCES

Gallagher, J. and G. Milkowski,1995: "Data Transport Within the Distributed Oceanographic Data System", 4th International World Wide Web Conference Proceedings, 691-700.